

# Curso Inteligência Artificial: do Zero ao Infinito

## Modelo MobileNet

Universidade Federal de Mato Grosso

# Agenda

- 1 Introdução
- 2 Spatial Separable Convolution
- 3 Depthwise Separable Convolutions

# Detecção de Objetos

## Introdução

- **R-CNN:** *Selective Search* + *AlexNet* + SVM.
- **Fast R-CNN:** *Selective Search* + *VGG16* + FC.
- **Faster R-CNN:** *RPN* + *VGG16* + FC

# Detecção de Objetos

## Introdução

	RCNN	Fast RCNN	Faster RCNN
<b>Test time per image with Proposals</b>	<b>50 seconds</b>	<b>2 seconds</b>	<b>0.2 seconds</b>
<b>(Speedup)</b>	<b>1x</b>	<b>25x</b>	<b>250x</b>
<b>mAP (PASCAL VOC 07)</b>	<b>66.0</b>	<b>66.9</b>	<b>66.9</b>

Fonte: Understanding Fast R-CNN and Faster R-CNN for Object Detection

# Detecção de Objetos

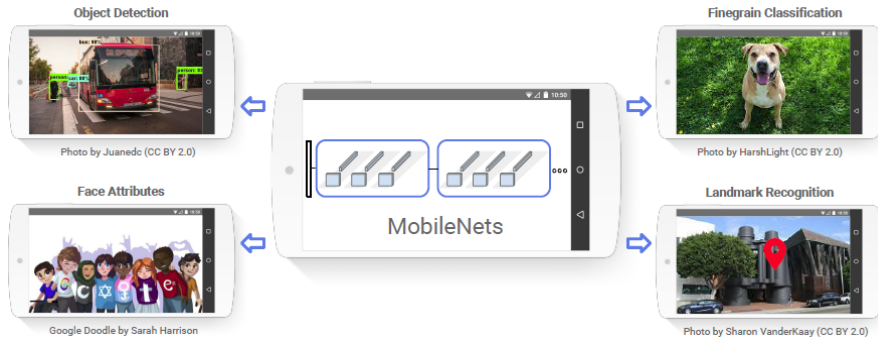
## MobileNet

- **MobileNet** é um modelo desenvolvido para ser usado em aplicações mobile.
- As convoluções foram substituídas por *Depthwise Separable Convolutions*.
- Reduziu o número de parâmetros em comparação com as convoluções regulares, pois requer um número menor de multiplicações e adições.

Fonte: MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications

# Detecção de Objetos

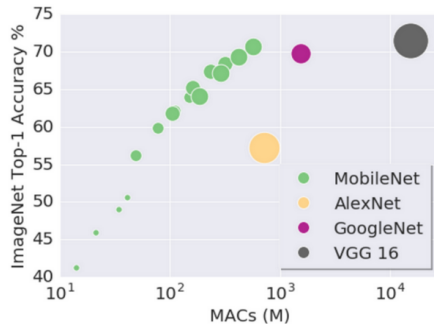
## MobileNet



Fonte: Image Classification With MobileNet

# Detecção de Objetos

## MobileNet



MACs (*Multiply-Accumulates*) é o número de operações de Multiplicação e Adição fundidas.

Fonte: Image Classification With MobileNet

# Agenda

- 1 Introdução
- 2 Spatial Separable Convolution**
- 3 Depthwise Separable Convolutions



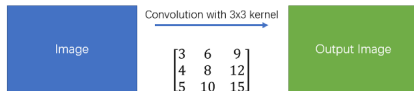
# Spatial Separable Convolution

- A *spatial separable convolution* é assim chamada porque manipula as dimensões espaciais de uma imagem e do kernel: a largura e a altura.
- Divide um kernel em dois kernels menores.
- Como exemplo, considere o filtro de Sobel:

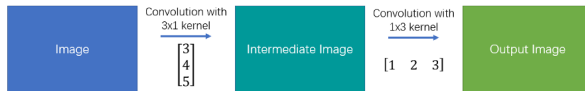
$$\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \times \begin{bmatrix} -1 & 0 & 1 \end{bmatrix}$$

# Spatial Separable Convolution

## Simple Convolution



## Spatial Separable Convolution



Fonte: A Basic Introduction to Separable Convolutions

# Spatial Separable Convolution

- Ao invés de fazer 1 convolução com 9 multiplicações, são feitas 2 convoluções com 3 multiplicações (6 no total) cada, que alcançam o mesmo efeito.
- Com menos multiplicações, a complexidade computacional diminui e a rede executa mais rapidamente.
- O principal problema é que com *spatial separable convolutions* nem todos os kernels podem ser “separados” em dois kernels menores.

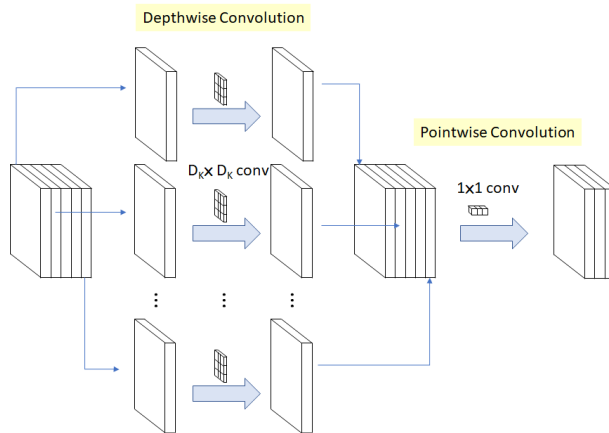
# Agenda

- 1 Introdução
- 2 Spatial Separable Convolution
- 3 Depthwise Separable Convolutions**

# Depthwise Separable Convolutions

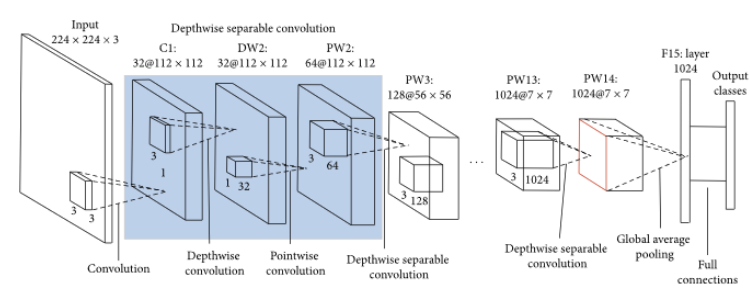
- *Depthwise separable convolutions* funcionam com kernels que não podem ser “fatorados” em dois kernels menores.
- Divide um kernel em 2 kernels separados que executam duas convoluções:
  - ▶ uma *depthwise convolution* e
  - ▶ uma *pointwise convolution*.

# Depthwise and Pointwise Convolutions



# Detecção de Objetos

## MobileNetV1



Fonte: Image Classification With MobileNet

# Detecção de Objetos

## MobileNetV1

Table 1. MobileNet Body Architecture

Type / Stride	Filter Shape	Input Size
Conv / s2	$3 \times 3 \times 3 \times 32$	$224 \times 224 \times 3$
Conv dw / s1	$3 \times 3 \times 32 \text{ dw}$	$112 \times 112 \times 32$
Conv / s1	$1 \times 1 \times 32 \times 64$	$112 \times 112 \times 32$
Conv dw / s2	$3 \times 3 \times 64 \text{ dw}$	$112 \times 112 \times 64$
Conv / s1	$1 \times 1 \times 64 \times 128$	$56 \times 56 \times 64$
Conv dw / s1	$3 \times 3 \times 128 \text{ dw}$	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 128$	$56 \times 56 \times 128$
Conv dw / s2	$3 \times 3 \times 128 \text{ dw}$	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 256$	$28 \times 28 \times 128$
Conv dw / s1	$3 \times 3 \times 256 \text{ dw}$	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 256$	$28 \times 28 \times 256$
Conv dw / s2	$3 \times 3 \times 256 \text{ dw}$	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 512$	$14 \times 14 \times 256$
5×	Conv dw / s1 $3 \times 3 \times 512 \text{ dw}$	$14 \times 14 \times 512$
	Conv / s1 $1 \times 1 \times 512 \times 512$	$14 \times 14 \times 512$
Conv dw / s2	$3 \times 3 \times 512 \text{ dw}$	$14 \times 14 \times 512$
Conv / s1	$1 \times 1 \times 512 \times 1024$	$7 \times 7 \times 512$
Conv dw / s2	$3 \times 3 \times 1024 \text{ dw}$	$7 \times 7 \times 1024$
Conv / s1	$1 \times 1 \times 1024 \times 1024$	$7 \times 7 \times 1024$
Avg Pool / s1	Pool $7 \times 7$	$7 \times 7 \times 1024$
FC / s1	$1024 \times 1000$	$1 \times 1 \times 1024$
Softmax / s1	Classifier	$1 \times 1 \times 1000$



- **MobileNet V2:** mudanças significativas foram feitas na arquitetura MobileNetV1 que resultaram em um aumento considerável na precisão do modelo.
- **MobileNet V3 Small:** é 6,6% mais precisa na classificação do que MobileNetV2 e tem latência semelhante.
- **MobileNet V3 Large:** é 3,2% mais precisa na classificação e reduz a latência em 20% quando comparado ao MobileNetV2.

Fonte: MobileNet V1 Architecture

# Referências

- MobileNet Architecture Explained
  - ▶ <https://prabinnepal.com/mobilenet-architecture-explained/>
- Review: MobileNetV1 - Depthwise Separable Convolution (Light Weight Model)
  - ▶ <https://towardsdatascience.com/review-mobilenetv1-depthwise-separable-convolution-light-weight-model-a382df364b69>
- MobileNet V1 Architecture
  - ▶ <https://iq.opengenus.org/mobilenet-v1-architecture/>

# Curso Inteligência Artificial: do Zero ao Infinito

## Modelo MobileNet

Universidade Federal de Mato Grosso